

#3

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant: Ernst F. Schröder  
Filed: Herewith  
For: VOICE CONTROL SYSTEM WITH A MICROPHONE ARRAY



**CLAIM OF PRIORITY UNDER 35 USC 119**

Hon. Assistant Commissioner for Patents  
Washington, D.C. 20231

Dear Sir:

Applicants hereby claim the priority under 35 USC 119 and under the International Convention for the Protection of Industrial Property, of German Patent Application Number 199 43 875.7 filed September 14, 1999. A copy of referenced patent application is enclosed herewith.

Respectfully submitted,  
Ernst F. Schröder

By:   
Kuniyuki Akiyama  
Patent Agent  
Reg. No. 43,314  
(609) 734-9404

THOMSON multimedia Licensing Inc.  
Two Independence Way  
P.O. Box 5312  
Princeton, New Jersey 08543

Date: September 12, 2000

# BUNDESREPUBLIK DEUTSCHLAND

Rest Available Copy  
Add'l copies available



## Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

**Aktenzeichen:** 199 43 875.7

**Anmeldetag:** 14. September 1999

**Anmelder/Inhaber:** Deutsche Thomson-Brandt GmbH,  
Villingen-Schwenningen/DE

**Bezeichnung:** System zur Sprachsteuerung mit einem  
Mikrofonarray

**IPC:** G 10 L 15/00

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Anmeldung.

München, den 06. Juli 2000  
Deutsches Patent- und Markenamt  
Der Präsident  
Im Auftrag

CERTIFIED COPY OF  
PRIORITY DOCUMENT

Hiebinger

## System zur Sprachsteuerung mit einem Mikrofonarray

Die Erfindung betrifft ein System zur Sprachsteuerung mit  
5 einem Mikrofonarray, die insbesondere zur Steuerung von  
Geräten der Unterhaltungselektronik genutzt werden kann.

### Stand der Technik

10

Sprachsteuerungssysteme finden in einer Vielzahl von  
technischen Gebieten Anwendung. Die gesprochenen Worte  
werden hierbei zunächst als Schallsignale detektiert,  
üblicherweise durch ein oder mehrere Mikrofone, und dann  
15 einem Spracherkennungssystem zugeführt. Die  
Spracherkennung basiert hierbei üblicherweise auf einem  
Akustik- und einem Sprachmodell. Das akustische Modell  
nutzt eine große Anzahl von Sprachmustern, wobei  
mathematische Algorithmen dazu verwendet werden, die  
20 akustisch am besten passenden Worte zu einem gesprochenen  
Wort anzugeben. Das Sprachmodell wiederum basiert auf  
einer Analyse, bei der anhand von einer Vielzahl von  
Dokumentproben festgestellt wird, in welchem Kontext und  
wie häufig gewisse Wörter normalerweise verwendet werden.  
25 Mit solchen Spracherkennungssystemen ist nicht nur das  
Erkennen einzelner Wörter, sondern auch von fließend  
gesprochenen Sätzen mit hohen Erkennungsraten möglich. Die  
Erkennungsrate sinkt jedoch drastisch, wenn nicht  
vernachlässigbare Hintergrundgeräusche vorliegen.

30

Die Robustheit gegen solche akustische Störeinflüsse kann  
auf verschiedene Weisen erhöht werden. So wird bei  
Diktiersystemen für Computer ein Mikrofon an einem

Kopfhörergestell direkt vor dem Mund des Sprechers befestigt. Bei diesen Systemen kann nur durch die unmittelbare Nähe zum Mund ein sehr konstantes Signal und damit eine zum Teil beachtliche Erkennungsrate erreicht werden. Ebenso ist es bekannt, ein Fernsehgerät zu steuern, indem die Bedienungsbefehle in das in einer Fernbedienung integrierte Mikrofon gesprochen werden. Auch hier muß jedoch die Fernbedienung unmittelbar vor den Mund des Benutzers gehalten werden.

10

### Erfindung

Der Erfindung liegt die Aufgabe zugrunde, ein System zur Sprachsteuerung anzugeben, welches eine ausreichende Störfestigkeit auch bei Spracheingabe aus größerer Ferne ermöglicht. Diese Aufgabe wird durch die in Anspruch 1 angegebene Vorrichtung gelöst.

Um eine Sprachsteuerung auch aus größerer Ferne zu ermöglichen, muß das Sprachsignal von störenden Hintergrundsignalen getrennt werden. Dies kann durch eine räumliche Trennung unter Verwendung von Mikrofonarrays aus zwei oder mehr Mikrofonen erfolgen. Hierbei ist es von Vorteil, die einzelnen Mikrofone des Mikrofonarrays räumlich über eine möglichst große Distanz zu verteilen. Bei einem einzelnen Unterhaltungselektronikgerät sind die Abstände zwischen den einzelnen Mikrofonen jedoch aufgrund der Geräteabmessungen begrenzt, wie z.B. bei einem Fernsehgerät auf unter einen Meter.

30

Im Prinzip besteht das erfindungsgemäße System zur Sprachsteuerung aus einem Mikrofonarray mit mehreren

Mikrofonen zur Umwandlung von Sprachbefehlen in elektrische Signale und einer zentralen Spracherkennungseinheit zur Umwandlung dieser elektrischen Signale in Bedienungsbefehle, wobei die Mikrofone auf  
5 verschiedene Geräte verteilt sind, die miteinander in solcher Weise verbunden sind, daß die durch die Mikrofone erzeugten Signale zu der zentralen Spracherkennungseinheit übertragen werden können.

10 Vorteilhafterweise werden hierbei die Geräte über ein bidirektionales Netzwerk verbunden, welches besonders vorteilhaft auf einem IEEE-1394-Bus basiert.

Besonders vorteilhaft ist das System, wenn ein oder  
15 mehrere Mikrofone in einem Unterhaltungselektronik-Wiedergabegerät, insbesondere einem Fernsehgerät, und ein oder mehrere weitere Mikrofone in externen Lautsprechern integriert sind.

20

### Zeichnung

Anhand der Zeichnung wird ein Ausführungsbeispiel der Erfindung beschrieben.

25

Diese zeigt eine erfindungsgemäße Anordnung zur Sprachsteuerung mit einem Mikrofonarray.

### 30 Ausführungsbeispiele

In Fig. 1 ist schematisch ein erfindungsgemäßes System dargestellt. An ein Fernsehgerät TV sind zwei externe

Lautsprecher LS1, LS2 angeschlossen. Interne, in der Figur nicht dargestellte, Lautsprecher des Fernsehgerätes ermöglichen zusammen mit den externen Lautsprechern eine Surround-Sound-Wiedergabe von Mehrkanaltonsignalen, z.B. die Wiedergabe von entsprechenden digitalen Audiosignalen gemäß dem MPEG-2- oder AC3-Standard. Die externen Lautsprecher sind hierbei über einen IEEE-1394-Bus, auch FireWire genannt, direkt an das Fernsehgerät angeschlossen, könnten aber ebenso an einen geeigneten Surround-Sound-Receiver angeschlossen werden. Die Verwendung eines IEEE-1394-Busses ist von Vorteil, da dieser eine schnelle Datenübertragung und die Kommunikation zwischen verschiedenen Geräten ermöglicht. Weiterhin kann bei Aktivlautsprechern gleichzeitig eine Stromversorgung über diesen Bus erfolgen.

Zur Detektion der Sprachsignale ist ein Mikrofonarray vorgesehen. Dieses besteht aus zwei in den Fernsehempfänger integrierten Mikrofonen MTV1 und MTV2 sowie je einem in die Lautsprechergehäuse integriertes Mikrofon MLS1 und MLS2. Diese setzen die detektierten Schallsignale in elektrische Signale um, die durch Verstärker verstärkt, durch AD-Wandler in digitale Signale umgewandelt und dann einer Signalverarbeitungseinheit zugeführt werden. Die Signale von den externen Lautsprechern werden hierbei ebenfalls über den IEEE-1394-Bus der Signalverarbeitungseinheit in dem Fernsehgerät zugeführt. Diese berücksichtigt den jeweiligen Aufenthaltsort des Benutzers durch eine unterschiedliche Skalierung bzw. Verarbeitung der detektierten Schallsignale. Weiterhin kann auch eine Korrektur der Mikrofonsignale bezüglich der von den Lautsprechern abgegebenen Schallsignale erfolgen. Das so bearbeitete

Signal wird dann einer Spracherkennungseinheit zugeführt, die die elektrischen Signale in Worte umwandelt. Die diesen Worten entsprechenden Befehle werden dann schließlich einem Systemmanager zur Steuerung des Systems zugeführt.

Die Skalierung bzw. Verarbeitung der detektierten Schallsignale durch die Signalverarbeitungseinheit erfordert, daß die räumliche Anordnung der Mikrofone bekannt ist. Diese ist für die in das TV-Gehäuse integrierten Mikrofone bereits herstellerseitig bekannt. Für die in den Lautsprechern angeordneten Mikrofone muß dagegen die Position relativ zu dem TV-Gerät noch ermittelt werden. Dieses kann durch ein Ausmessen und eine Eingabe der ermittelten Werte über ein entsprechendes Bildschirmmenü erfolgen. Ebenso kann jedoch ein Mess- und Abgleichvorgang durchgeführt werden, indem durch die Lautsprecher ein Testsignalton wiedergegeben und von den Mikrofonen detektiert wird und aus den unterschiedlichen Laufzeiten die Position der Mikrofone ermittelt wird.

Die beiden in dem Fernsehempfänger integrierten Mikrofone können vorteilhaft auf der linken und rechten Seite des Gehäuses des jeweiligen Gerätes untergebracht werden. Diese Mikrofonanzahl und ebenso die Anzahl der weiteren Mikrofone ist jedoch keinesfalls auf zwei beschränkt. Ebenso sind eine Vielzahl von Kombinationen denkbar, in denen die Mikrofone integriert sind. So können Mikrofone statt oder zusätzlich zu dem Fernsehgerät auch in einem Videorekorder, DVD-Spieler oder einer Fernbedienung integriert werden. Selbst eine Installation von Mikrofonen in Geräten, die sich in verschiedenen Räumen befinden, ist möglich.

Weiterhin ist die Verbindung der Geräte nicht auf ein Bussystem beschränkt. So ist es z.B. denkbar, bei der Ausführungsform aus Figur 1 die Lautsprecher stattdessen  
5 mit Funksignalen anzusteuern. In diesem Fall müssen die einzelnen Lautsprecher jedoch zusätzlich zu einem Funkempfänger auch einen Funksender aufweisen.

10 Die Erfindung kann zur Sprachfernbedienung von verschiedensten Geräten der Unterhaltungselektronik, wie z. B. von TV-Geräten, Videorecordern, DVD-Spielern, Satellitenempfängern, TV-Video-Kombinationen, Audiogeräten oder kompletten Audiosystemen, aber ebenso von  
15 Personalcomputern oder von Haushaltsgeräten eingesetzt werden.



### Patentansprüche

1. System zur Sprachsteuerung mit einem  
5 Mikrofonarray bestehend aus mehreren Mikrofonen (MTV1, MTV2, MLS1, MLS2) zur Umwandlung von Sprachbefehlen in elektrische Signale und mit einer zentralen Spracherkennungseinheit zur Umwandlung dieser elektrischen Signale in Bedienungsbefehle, **dadurch**  
10 **gekennzeichnet**, daß die Mikrofone auf verschiedene Geräte (TV, LS1, LS2) verteilt sind, die miteinander in solcher Weise verbunden sind, daß die durch die Mikrofone erzeugten Signale zu der zentralen Spracherkennungseinheit übertragen werden können.

15

2. System nach Anspruch 1, **dadurch gekennzeichnet**, daß die Geräte über ein bidirektionales Netzwerk verbunden sind.

20

3. System nach Anspruch 2, **dadurch gekennzeichnet**, daß das bidirektionale Netzwerk auf einem IEEE-1394-Bus basiert.

25

4. System nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, daß ein oder mehrere Mikrofone in einem Unterhaltungselektronik-Wiedergabegerät (TV), insbesondere Fernsehgerät, und ein oder mehrere weitere Mikrofone in externen Lautsprechern (LS1, LS2) integriert sind.

### Zusammenfassung

Sprachsteuerungssysteme finden in einer Vielfalt von  
5 technischen Gebieten Anwendung. Die gesprochenen Worte  
werden hierbei durch ein oder mehrere Mikrofone detektiert  
und dann einem Spracherkennungssystem zugeführt. Um eine  
Sprachsteuerung auch aus größerer Ferne zu ermöglichen,  
muß das Sprachsignal von störenden Hintergrundsignalen  
10 getrennt werden. Dies kann durch eine räumliche Trennung  
unter Verwendung von Mikrofonarrays aus zwei oder mehr  
Mikrofonen erfolgen. Hierbei ist es von Vorteil, die  
einzelnen Mikrofone des Mikrofonarrays räumlich über eine  
möglichst große Distanz zu verteilen. Bei einem einzelnen  
15 Unterhaltungselektronikgerät sind die Abstände zwischen  
den einzelnen Mikrofonen jedoch aufgrund der  
Geräteabmessungen begrenzt. Das erfindungsgemäße System  
zur Sprachsteuerung besteht daher aus einem Mikrofonarray  
mit mehreren Mikrofonen (MTV1, MTV2, MLS1, MLS2), die auf  
20 verschiedene Geräte (TV, LS1, LS2) verteilt sind, wobei  
die durch die Mikrofone erzeugten Signale zu der  
zentralen Spracherkennungseinheit übertragen werden  
können, vorteilhafterweise über ein bidirektionales  
Netzwerk basierend auf einem IEEE-1394-Bus.

25

Fig. 1

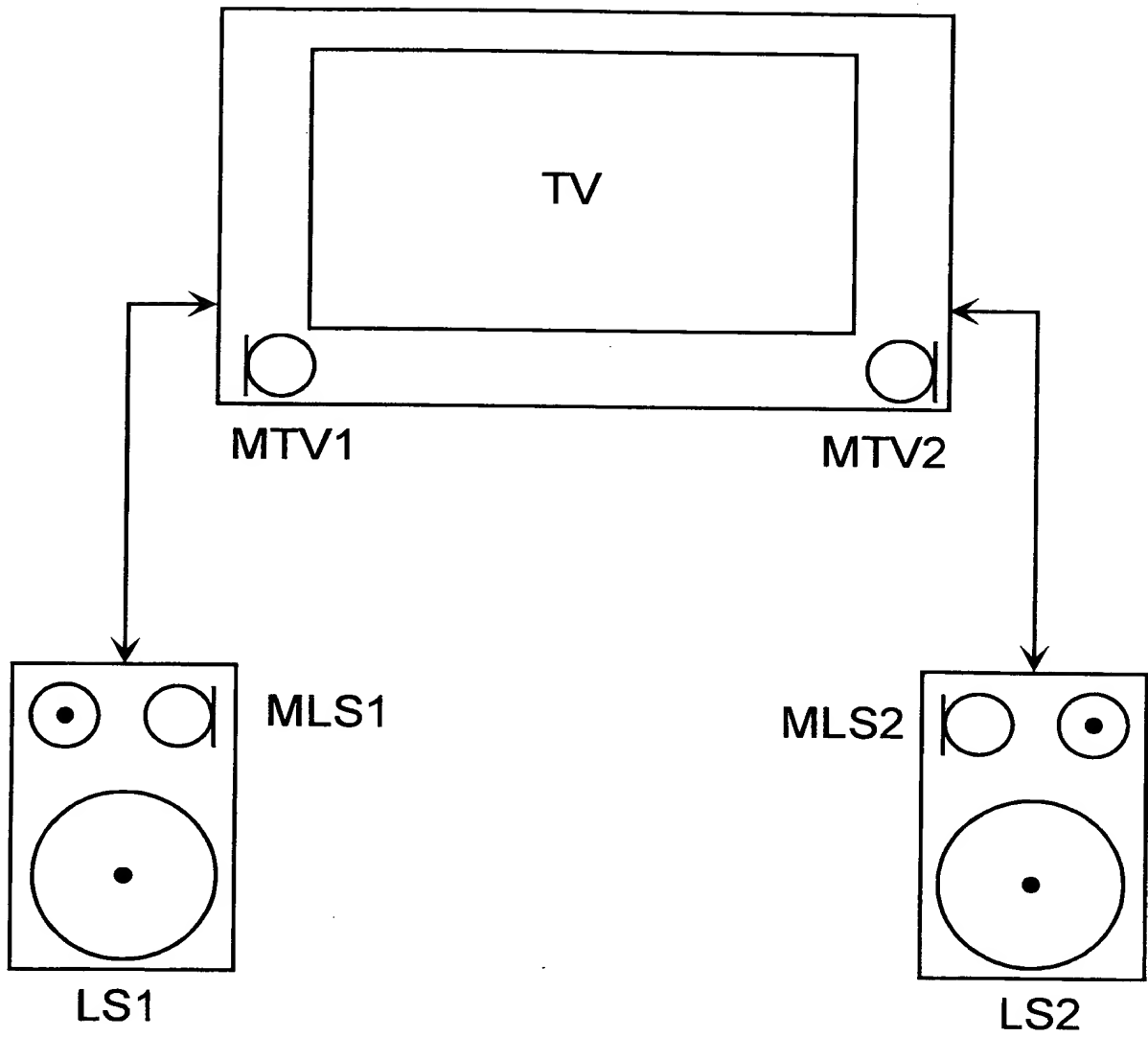


Fig. 1